

MAPS Outlier Output Guide

The detailed output from a Ratio Outlier List Search in MAPS provides several columns of information. This guide is designed to explain the output and how to interpret it. The columns are as follows:

Standard output

1. GenBank Accession number
2. GeneID: IMAGE clone ID for human & mouse cDNA, accession number for human oligo and rat.
3. Replicate Exp Hit Ratio: ratio of the number of arrays in which the clone was differentially expressed to the total number of replicate arrays that were done.
4. Flag: identifies clones that have been flagged for certain functions or pathways.
5. Description: current name assigned to the clone in the Unigene database.
6. Unigene cluster: cluster ID corresponding to the name given in the previous column.
7. Gene Symbol
8. Chromosome Map
9. Transcription Map
10. \log_2 Cal. Ratios: calibrated ratio value for each hybridization in which the clone was differentially expressed.
11. HybID: internal MAPS identifier for a chip.
12. $\log_2(Ch1 * Ch2)$: a measure of overall spot intensity; a cutoff can be set on the Outlier Search page to flag observations having a total intensity below the cutoff.
13. Ch1/Ch2: gives the mean background-subtracted intensities for both channels.
14. Modified z-scores: this statistic is calculated to identify spots in a set of replicates that are outliers relative to the others. For a given log ratio value x , the score is calculated as

$$Z = \frac{0.6745(x - \tilde{x})}{MAD}$$

where $MAD = \text{median} |x - \tilde{x}|$ and \tilde{x} is the median log calibrated ratio. A cutoff can be specified on the outlier search page; spots with modified Z-scores exceeding this threshold will be indicated in red in the results table. In general, large Z scores indicate observations that deviate significantly from the median.

15. Mean: the mean \log_2 calibrated ratio, \bar{x} .
16. Modified mean: the mean \log_2 ratio value calculated after deleting observations having modified Z-scores above the specified cutoff.
17. Median: median of the set of \log_2 calibrated ratios, \tilde{x} .
18. Standard deviation: measure of the spread of the n \log_2 calibrated ratio values about their mean, calculated as

$$s = \sqrt{\frac{1}{n-1} \sum_i^n (x_i - \bar{x})^2}$$

This estimate is not particularly good for small values of n .

19. Standard error: measure of the variation of the *mean* \log_2 calibrated ratio value, calculated as

$$SE(\bar{x}) = \frac{s}{\sqrt{n}}$$

20. CV: the *coefficient of variation*, calculated as

$$CV = \frac{s}{\bar{x}}$$

Optional output

- PRC - the plate number, row, and column from which the given clone was printed onto the array. This information is necessary for slides having multiple spots of the same clone, since GeneID alone will not differentiate these spots.